Building a GWA Platform in Maize with Illumina Technology

Joseph D. Clarke – Integrated Genome Biology
Syngenta Biotechnology, Inc., RTP, NC

Kishore Venkata & Daolong Wang – Marker-Trait Association Genetics
Syngenta Seeds Research, Clinton, IL

Illumina Agricultural Genomics Seminar Series, RTP, NC **August 6, 2009**

About Syngenta

- Leading global agribusiness
- Strong worldwide presence
- ~24,000 employees in over 90 countries
- Rich pipeline of Crop Protection and Seeds products

Syngenta Biotechnology Inc.,

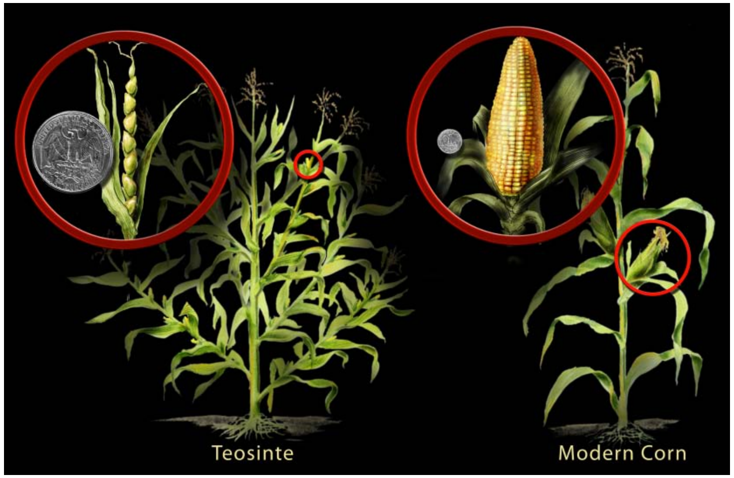
- Formed in RTP, NC in 1984
- Strong ties with the community and local academic institutions
- State-of-the-art facilities
  - 120,000 sq ft of lab space
  - 50,000 sq ft of greenhouse space
  - 27,000 sq ft of office space
- Around 400 employees
- Expansion is underway

Agenda

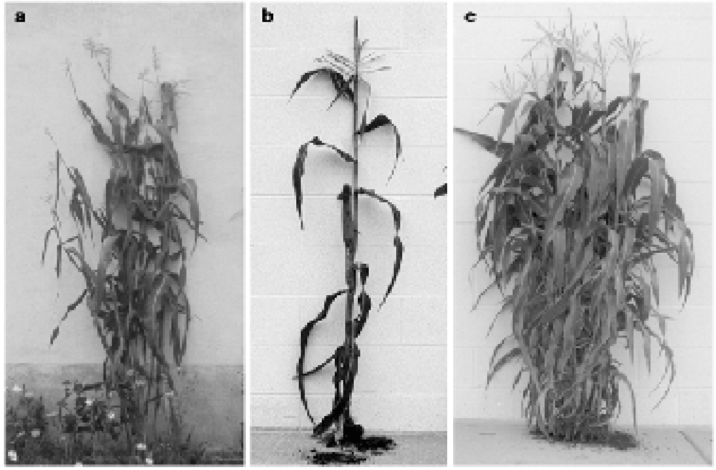
- Building a GWA Platform in Maize with Illumina Technologies
  - About Maize and Association Mapping
  - Project Objectives
  - Data Generation
  - Data Analysis: SNPs and Coverage
  - Considerations for GWA in Maize with Infinium

Diversity in Maize

- The genetic diversity of maize is greater than the diversity between human and chimps
  - 18 genes in 36 lines (Ching et al., 2002)
    - ~1 SNP per 31 bp in non-coding regions
    - ~1 SNP per 124 bp in coding regions.
    - Insertions and deletions are frequent in non-coding regions (1 per 85 bp), but rare in coding regions.
  - Would need to self maize for 4 generations to reduce diversity to human level (Tenallion et al., 2001).



Nicolle Rager Fuller, National Science Foundation

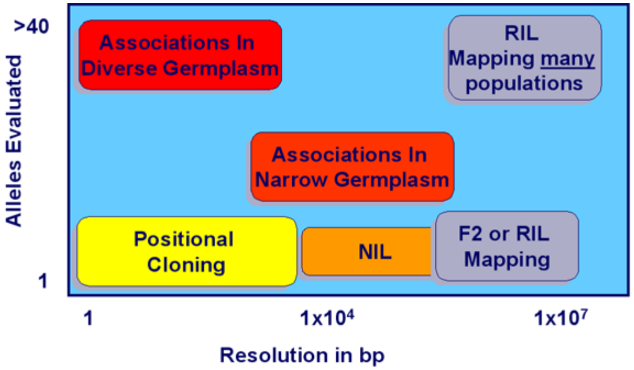


Doebley et al., 1997

Nature Reviews | Genetics

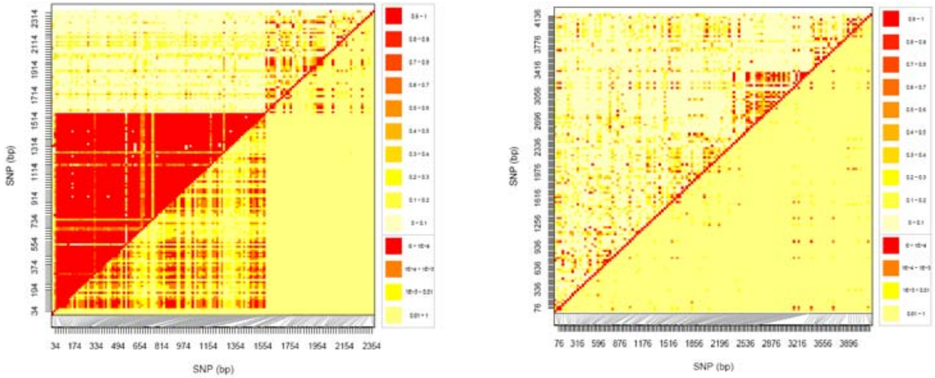
Association Mapping in Maize

- Association studies assess correlations between genetic variants and trait differences on a population scale



Modified from Buckler et al., 2008

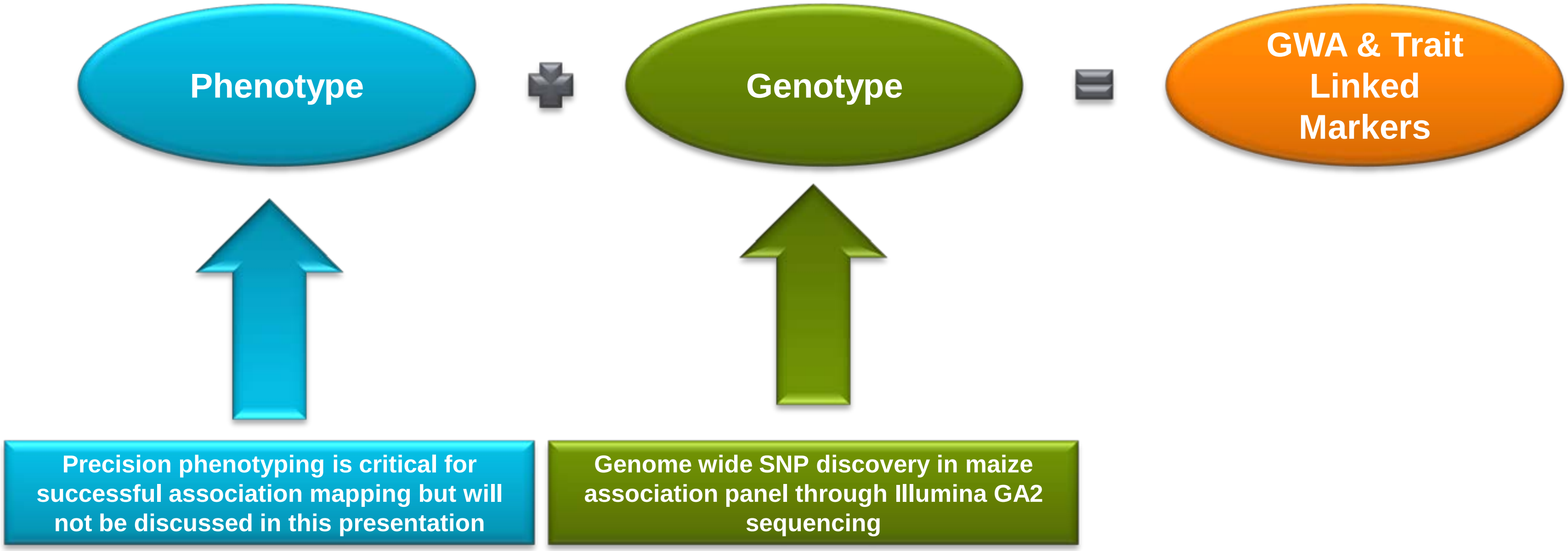
- Strongly dependent on Linkage Disequilibrium: the nonrandom association of linked alleles



LD decay between 2 random genes across a maize population
Briggs, B., et al., Syngenta

Genome Wide Association Requires Lots of SNPs

- Project Plan
  - **SNP discovery and genotyping of Maize Association Panel**
  - Algorithm and Statistical method development
  - Application of Genome Wide Association capability



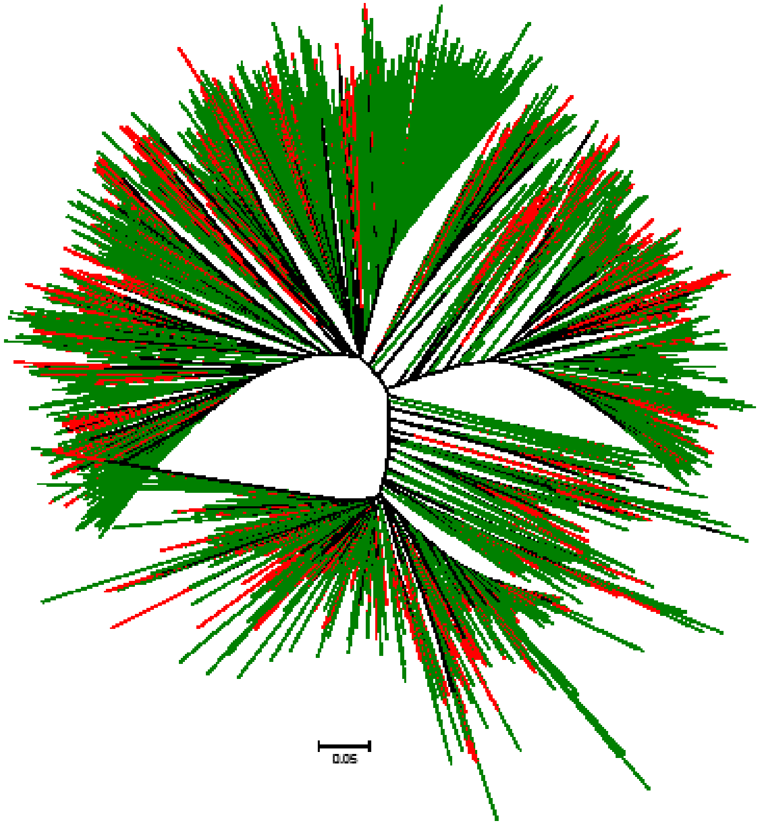
Phenotype + Genotype = GWA & Trait Linked Markers

Precision phenotyping is critical for successful association mapping but will not be discussed in this presentation

Genome wide SNP discovery in maize association panel through Illumina GA2 sequencing

Selecting the Appropriate Germplasm

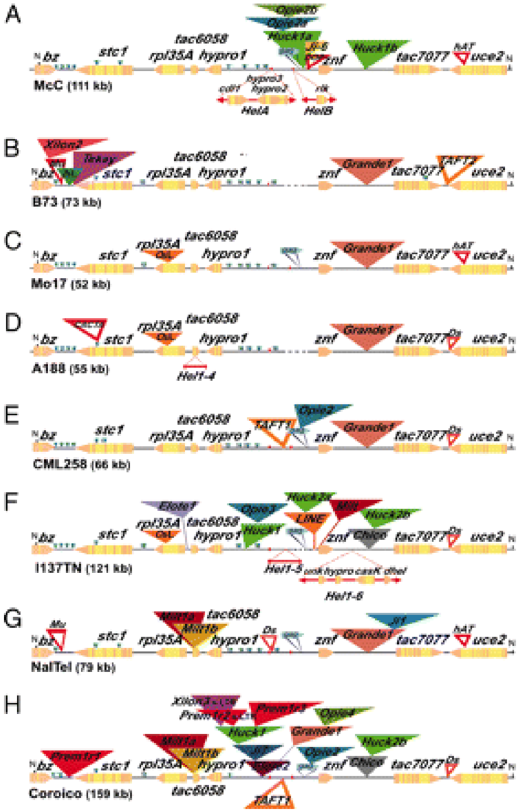
- The Maize Association (sub)Panel was selected by Syngenta association geneticists based on optimal genotypic diversity representing a broad collection of maize inbreds selected for their phenotypic and regional diversity and includes additional lines to broaden panel utility.



Genotypic diversity and relatedness across maize lines
Guiterez, A., et al, Syngenta

Selecting the Appropriate Method

- The success of Genome Wide Association depends on ensuring common sequence from multiple lines in the panel is represented. The Syngenta Project team decided Transcriptome sequencing best fit the criteria.



Wang Q., et al., 2006

Partnership Between Syngenta and NCGR Scientists

- Partnership with the National Center for Genome Resources (NCGR)
  - 650 Plant lines grown and sampled at SBI
  - 102 GA2 runs performed at NCGR
  - 160 GB of quality sequence obtained for 650 lines averaging ~247 MB per line

Overcoming Technical Challenges with NCGR

- Building a high throughput cDNA conversion and GA2 library prep protocol

RNA Area: 1,601.9

RNA Concentration: 1,005 ng/µl

rRNA Ratio [28s / 18s]: 1.3

RNA Integrity Number (RIN): N/A (B.02.05)

Result Flagging Color:

Result Flagging Label: RIN N/A

NCGR High Throughput mRNAseq Protocol Development

Building a Reference Genome and Quality Criteria

- Developing a Maize genome reference data set by masking overlapping sequence

Masked due to overlap with another BAC

Masked due to repetitive elements

Exposed for alignment and analysis

Rose, M., et al., Syngenta

- Developing quality criteria and formats to interpret data:
  - Validation criteria to ensure >90% confidence for low read variant calls
  - Genotyping categories to better describe data for downstream analysis
  - Capturing sequence coverage for downstream analysis (with NCGR)

Skalla, D., et al., Syngenta with NCGR

SNP Discovery in Maize Association Panel

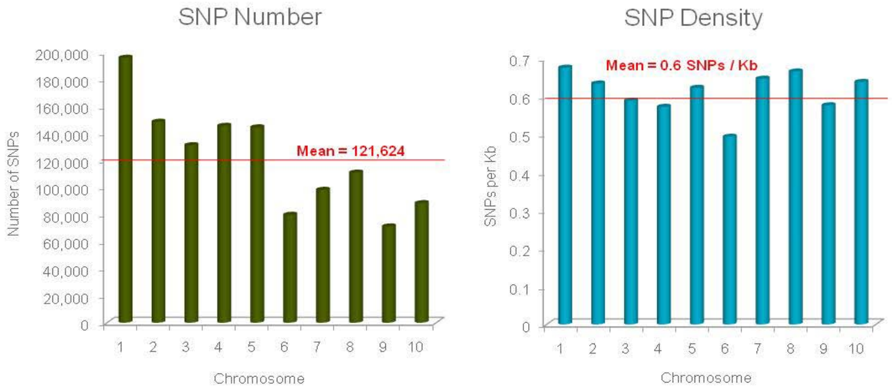
- Represents polymorphic and variant information across the data set
  - Difficult to accurately account for variants and non-variants while imposing filtering criteria

- Estimated quality SNPs discovered
  - 9.5M

- Estimated genotyping information as a function of line number
  - 100% ~4,500 genotyped positions
  - 90% ~650,000 genotyped positions
  - 80% ~1.8M genotyped positions
  - 70% ~3.3M genotyped positions

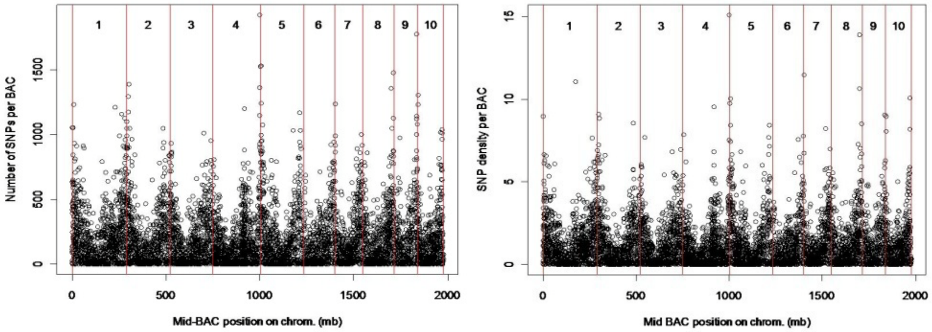
Determining Genome Distribution of SNPs

- Number of SNPs and average SNP density per chromosome



Wang, D., et al., Syngenta

- Number of SNPs and SNP density by BAC location across chromosome



Wang, D., et al., Syngenta

Conclusions on GA2 SNPs for Genome Wide Association

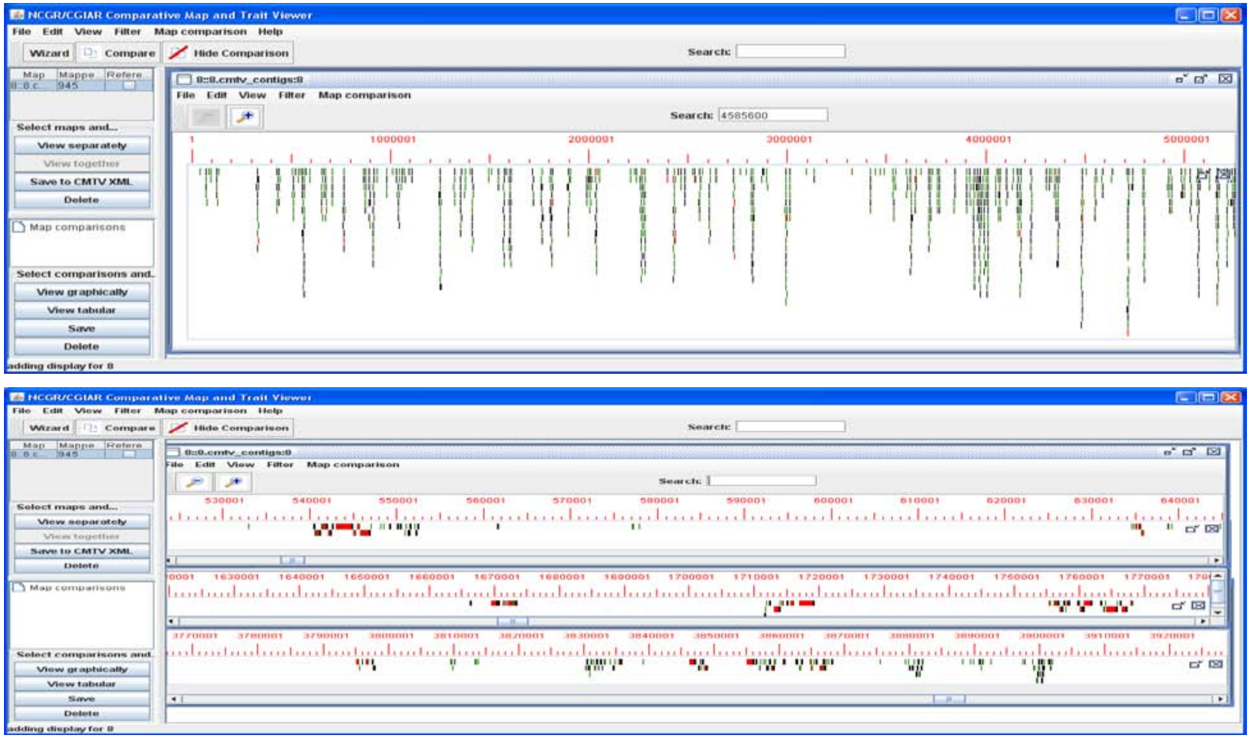
- Considering
    - High SNP number and quality validation criteria
    - Gene specific diversity captured and haplotype preservation
    - ~20-30K genes tagged with sequence in each line
    - Range of rare to common alleles identified
    - Sequence distribution across maize genome is random and well represented
    - Common regions of sequence coverage across maize panel identifiable
    - LD calculations possible for subset of genes with sufficient sequence coverage across panel

- Conclude the experiment is a cost effective design to build a competitive GWA platform that is trait and population independent.

- GWA analyses for multiple traits currently in progress.

Beyond SNP Discovery for GWA

- New gene discovery and annotation
- Genome architecture
- Allele specific gene expression



Transcript coverage across 5 MB and 3 x 100KB of the maize genome (NCGR)

Acknowledgements

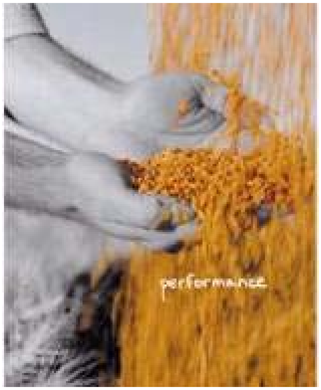
Syngenta

Shannon McDonald
Dale Skalla
Bob Dietrich
Todd Moughamer
Sean Korb
Daolong Wang
Bill Briggs
Joe Byrum
Lynn Senior
Joe Clarke

Vance Kramer
Sarah Muncie
Molly Dunn
Mark Rose
Kishore Venkata
Todd Warner
Andres Gutierez
Chris Zinselmeier
Erik Legg

NCGR

Greg May
Andrew Farmer
Ryan Kim
Raymond Langley
Jimmy Woodward
Leonda Clendenen
Neil Miller



Thank you